Government of Western Australia
Department of Health

Classification: Official

WA Health
Data Linkage Services

Limitations and Suitable Use of Linked Data

24 December 2025

Version Control and Approval

This document should be considered a 'live document' and is reviewed regularly and updated as required to:

- Reflect changes to policy and/or procedures
- Incorporate stakeholder feedback
- Determine effectiveness, relevance, and currency

Review and update of this document is coordinated by the Data and Information Systems unit within the Information and System Performance Directorate (ISPD).

Version	Published date	Approved by	Amendment(s)
1.0	08 January 2024	Pavi Krishnan, A/Manager	Original Version
1.1	24 December 2025	Russell Miln, Manager	Added details on factors influencing data quality Included explanation of false positives and false negatives Simplified content by removing unnecessary detail and focusing on high-level guidance for population/cohort analysis

Contact

Enquiries relating to this guide may be directed to:

Title:	Manager, Data Linkage Services
Directorate:	Information and System Performance
Email:	DataLinkageStrategy@health.wa.gov.au

Table of Contents

1. Introduction

1.1 Background

Data Linkage is a technique for establishing unique identifiers (data linkage keys) which are used to connect information from different data sources thought to relate to the same person, family, place or event via data integration processes. Information is created when a person encounters certain services, for example, when they visit an emergency department, stay in a hospital or register the birth of their child.

The Data Linkage Services branch at the Department of Health uses probabilistic linkage techniques to link data in WA. As datasets from various sources do not share a universal ID, records must be linked probabilistically to support data integration processes. The data linkage process is complex and primarily utilises a range of demographic data variables (or otherwise known as ‘reasonably identifiable information’ such as name, date of birth and address) to bring together service records that likely belong to the same individual, place, or event, generally utilising a software-driven probabilistic matching process.

The WA Data Linkage System (WADLS), managed and operated by the Information and System Performance Directorate within the Department of Health is used to create, store, update and extract links between datasets. Please visit the Data Linkage Services website for more information on the complex data linkage process.

1.2 Purpose

The purpose of this document is to articulate the limitations to and the appropriate use of linked data.

2. What are the limitations to linked data?

While data linkage techniques allow records to be linked where it would not otherwise be possible, it is important to consider the limitations caused by the probabilistic nature of linkage.

- Linkage quality depends heavily on the quality of demographic information in each dataset. The consistency and completeness of identifiers vary widely across collections. Common data quality issues include:
    - Data entry errors, such as typographical mistakes or miscommunication when information is collected verbally.
    - Missing information (e.g., date of birth) which can make it difficult to correctly link records, particularly when names are common.
    - Contradictory information which increases the risk of mismatches (e.g., a child's date of birth recorded with a parent's name, or two records with the same Medicare number but different names and dates of birth).
    - Corrections made between data submissions, where updated information may not align with earlier extracts used for linkage.
    - Assumptions made about certain data items, such as treating an address as a permanent residential address where it may be temporary.
- Due to the probabilistic nature of linkage, there is a potential for instances of 'false positives' where records are erroneously linked, as well as the possibility of 'false negatives' where records that should be linked are missed.
    - The former (false positives) may include records that have been corrected by the collection since they were originally received (e.g., errors identified and corrected after the data has been provided to linkage, but a link has already been established).
    - The latter (false negatives) may include records that are deliberately not linked to maintain the integrity of the WADLS. Data linkage is an iterative process and links are regularly created, modified, and deleted as new information becomes available.
- Each dataset represents only a subset of the WA population, based on the purpose and circumstances of the collection. For example, a given data collection may only include records from public, not private health providers, meaning the linked data may not represent the full population.

3. How should linked data be used?

The Data Linkage Services branch employs a variety of strategies to promote high quality linkage. However, ‘perfect’ data quality and linkage cannot be guaranteed. Therefore, **linked data should not be used at the level of individual records**; that is, linked data should not be used to seek, interpret or act upon information about any one individual, unless deemed appropriate by the Data Steward or Data Custodian of the linked information.

Linked data is best suited for population-level, or cohort-level analysis, such as identifying trends over time, monitoring service use, or evaluating the impact of programs and policies.

Users should ensure they understand any known limitations or characteristics of the datasets they are analysing. Researchers and Department of Health analysts can do this by consulting the relevant Data Custodian and reviewing supporting materials such as data validation reports, data specifications, and key performance indicator definitions.

As noted above, data linkage is not perfect, therefore users of linked data should always undertake standard data cleaning to enhance accuracy, reliability and clarity.

This document can be made available in alternative formats on request for a person with disability.

health.wa.gov.au